January 13, 2024

Hon. Ona T. Wang
United States Magistrate Judge
Southern District of New York

> Re: *The New York Times Company v. Microsoft Corp.* No.: 23-cv-11195:
>   Dispute Regarding Microsoft's Responses to Document Requests

Dear Magistrate Judge Wang:

Plaintiff The New York Times Company ("The Times") seeks an order compelling Microsoft to produce the categories of documents described below.[1]

1. Documents Concerning the Market for Training Data and RAG Data

The Times seeks an order requiring Microsoft to produce internal and external communications concerning Microsoft's negotiations about content licensing agreements, including for training models and for use with RAG,[2] in response to RFPs 61, 113, and 114.[3]

These requested documents and communications are relevant to The Times's allegation that there is a market for the at-issue data. FAC ¶ 155; *see A&M Recs., Inc. v. Napster, Inc.*, 239 F.3d 1004, 1016 (9th Cir. 2001) ("Fair use, when properly applied, is limited to copying by others which does not materially impair the marketability of the work which is copied." (quoting *Harper & Row Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 566-67 (1985)). The Times reasonably seeks documents and communications concerning Microsoft's efforts to acquire content for model training and RAG, for use in its generative AI products such as Microsoft Copilot.

Like OpenAI, Microsoft acknowledges this information is relevant, and has agreed to produce some of the requested documents – specifically, executed content licensing agreements, and internal and external communications related to those executed agreements. But Microsoft refuses to produce communications about licensing negotiations that did *not* result in a final executed agreement, or any documents about the existence or non-existence of a licensing market.

---

[1] The parties met and conferred about these disputes by video conference on January 8 and 9. While the parties resolved some of their disputes, they were unable to resolve the disputes addressed in this motion. The Times's Fourth Set of RFPs (Requests 92-129) and Microsoft's Responses are attached hereto as Exhibits 1 and 2. *See also* Dkt. 381-1 (email correspondence between the parties).

[2] RAG enables Defendants' products to integrate content from the "live" web—including The Times's website—with their large language models. By using RAG, Defendants' products can generate answers to queries about current events and other information that postdates the training of the models. FAC ¶¶ 81, 108-23, 163, 179.

[3] RFP 61 seeks "Documents and communications concerning negotiations between Defendants and owners of copyrighted content, including The Times, over Defendants' use of copyrighted content in Defendants' Generative AI Models and Defendants' Generative AI Products or Services." RFP 113 seeks "Documents concerning the market for training data, including any discussion of the existence or non-existence of such a market and the actual or potential monetization of training data." RFP 114 seeks "Documents concerning the market for content for retrieval-augmented generation or generative search features, including any discussion of the existence or non-existence of such a market."

With respect to negotiation documents related to potential deals, these communications can be just as probative of the existence of a market for content as negotiations that did result in an agreement. For example, if communications show that Microsoft was willing to pay a news organization a certain amount of money to use their content for Microsoft Copilot, that admission would be relevant to assessing the market for news data, even if that deal never materialized.

Microsoft contends this Court's order in the related *Authors Guild* case forecloses The Times's request for these communications. *See Authors Guild*, Case No. 23-cv-8292, Dkt. 293 at 3 (denying the class plaintiffs' motion filed at Dkt. 269). But that order is distinguishable, because there Microsoft opposed producing communications about potential licensing agreements on the basis that the class plaintiffs were seeking communications about licensing agreements for models and products not at issue in that case. *See* Dkt. 279 at 1 (arguing that "because Microsoft's LLMs are irrelevant, so are documents related to unexecuted data access agreements"). Microsoft made the same argument at the ensuing conference, stating that while the class case is "focused on the training of *OpenAI's* LLMs," the class plaintiffs were improperly seeking documents about "negotiations regarding data access for *Microsoft's* LLMs," thereby implicating "an entirely different set of LLMs that are not at issue." *Id.* at 138:2-139:7 (emphases added).

The Times's case, by contrast, includes claims concerning Microsoft Copilot, a user-facing chatbot that relies on an OpenAI large language model and provides responses to user queries, including by integrating content from the live web. FAC ¶¶ 81, 108-23. Microsoft's licensing agreements (and related negotiations) are therefore relevant to this case. Indeed, Microsoft has previously distinguished the News Cases from the class actions by claiming that Copilot is only relevant to the News cases. *See* Dkt. 148 at 4 (asserting that the News cases "concern Microsoft's generative AI products (such as Bing Copilot) that consist of underlying generative AI models as well as, and in combination with, innovative search engine tools that allow a user to enter any prompt," and that "Microsoft's generative search technology is unique to the Newspaper Cases and does not feature in the Consolidated Class Actions"). Because The Times seeks licensing communications about products that have been indisputably subject to discovery in this case (like Microsoft Copilot), the *Authors Guild* ruling does not apply.

Finally, in response to RFPs 113 and 114, Microsoft should produce documents that address the existence of a market for the data, regardless of whether such documents address a particular negotiation with a particular partner. These requests go to the heart of the fourth factor of the fair use analysis, "the effect of [Defendants'] use upon the potential market for or value of [The Times's] copyrighted work." 17 U.S.C. § 107. As the Second Circuit recently explained, this factor is "undoubtedly the single most important element of fair use." *Hachette Book Grp., Inc. v. Internet Archive*, 115 F.4th 163, 189 (2d Cir. 2024). And while The Times can easily show that there is a market for its licensed work generally, proving that there is a licensing market for training and RAG will further demonstrate that Defendants' unlicensed use "usurp[s] the market for the original[] by offering a competing substitute." *Id.* at 190. There is simply no excuse for withholding these plainly relevant documents.

2. Documents Concerning Microsoft's Investment in OpenAI

The Times seeks an order requiring Microsoft to produce documents concerning its investments in OpenAI. Specifically, RFP 123 seeks "Documents concerning whether Microsoft's potential or actual investments in OpenAI, including in the years 2016, 2019, and 2024, would

affect OpenAI's non-profit status, mission, or ability to engage in non-commercial activities." These documents are relevant because the commercial nature of Defendants' products is a key component of the fair use inquiry. See 17 U.S.C. § 107(1) (requiring analysis of "the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes").

Microsoft has agreed to produce some documents in response to this request, but has objected to providing discovery about its 2024 investments on the grounds that The Times's case was filed in December 2023. But the News Plaintiffs allege ongoing misconduct, requiring discovery into the ongoing commercial exploitation of their content. *See, e.g.*, Times FAC ¶ 167 (describing "Defendants' ongoing infringing conduct"); *id.* ¶ 66 (alleging that Microsoft "continues to" collaborate with OpenAI to infringe Times works). "Rule 26 does not limit discovery to pre-complaint events in cases such as this in which the complaint alleges that [the wrongful conduct] was continuing as of the date the complaint was filed." *Robco Distributors, Inc. v. Gen. Glass Int'l Corp.*, 101 F.R.D. 547, 548 (E.D.N.Y. 1984); *see also Deluxe Fin. Servs., LLC v. Shaw*, 2017 WL 7369890, at *5 (D. Minn. Feb. 13, 2017) ("it may be appropriate to allow post-complaint discovery where documents requested are relevant to the claims and defenses in the case, especially where the complaint alleges continuing violations").[4] Microsoft's ongoing investments in OpenAI are plainly relevant to The Times's contributory and vicarious liability claims against Microsoft, which of course raise questions about Microsoft's contributions to and investments in OpenAI. FAC ¶¶ 66, 151, 169. And very recent documents are especially likely to be relevant, as OpenAI announced in December 2024 that it would undergo a significant corporate restructuring, including to convert its for-profit arm into a public benefit corporation, citing the need to "consider the interests of those who would finance the mission" (i.e., investors like Microsoft).[5]

3. Documents Concerning the Impact of Microsoft's Conduct on Content Creators

Finally, The Times seeks an order requiring Microsoft to produce documents in response to RFP 109, which seeks "Documents reflecting the actual or expected financial impact of Defendants' Generative AI Products and Services on copyright owners, including News Plaintiffs and other content creators." Microsoft has agreed to produce some documents in response to this request but is limiting its investigation to documents addressing the specific three News Plaintiffs. As a compromise, The Times has offered to narrow this request to the actual or expected financial impact on the publishing industry, the news industry, cooking websites, and product recommendation websites, including the News Plaintiffs, but Microsoft has still refused to search for these documents (although it says it will not withhold them should it encounter them by chance). These documents are relevant to the fourth factor of the fair use analysis, as they go to the market impact of Defendants' misappropriation of copyrighted works. See p. 2, *supra* (citing *Hachette*). This Court should order Microsoft to produce documents in response to this RFP, as narrowed by The Times's compromise proposal.

---

[4] Even if it were appropriate to cut off discovery as of the date of filing the Complaint, the consolidated Daily News and CIR cases were filed in April and June 2024, respectively, and thus should encompass 2024 documents.

[5] https://openai.com/index/why-our-structure-must-evolve-to-advance-our-mission/.

Respectfully,

/s/ Ian B. Crosby
Ian B. Crosby
Susman Godfrey L.L.P.

cc: All Counsel of Record (via ECF)